$770.00 -101 $44.00-103

INTELLECTUAL PROPERTY LAW

KNOBBE, MARTENS, OLSON & BEAR

A LIMITED LIABILITY PARTNERSHIP INCLUDING
PROFESSIONAL CORPORATIONS

PATENT, TRADEMARK AND COPYRIGHT CAUSES

KOLL CENTER SAN DIEGO

501 WEST BROADWAY

SUITE 1400

SAN DIEGO, CALIFORNIA 92101-3505

(619) 235-8550

FAX (619) 235-0176

INTERNET: WWW.KMOB.COM

KNOBBE*
MARTENS*†
OLSON*†
BEAR
OLSON*
BUNKER
MICHAELSEN
S. HAMILTON
WILLIAM H. NIEMAN
LOWELL ANDERSON
ARTHUR S. ROSE†
JAMES F. LESNIAK
JERRY T. SEWELL
JOHN B. SGANGA, JR
EDWARD A. SCHLATTER
GERARD VON HOFFMANN
JOSEPH R. RE
CATHERINE J. HOLLAND
JOHN M. CARSON
KAREN VOGEL WEIL†
ANDREW H SIMPSON
JEFFREY L. VAN HOOSEAR
DANIEL E. ALTMAN
ANITA M. KIRKPATRICK
ERNEST A BEUTLER

MARGUERITE L GUNN
STEPHEN C. JENSEN
VITO A. CANUSO III
WILLIAM H. SHREVE
LYNDA J. ZADRA-SYMES††
STEVEN J. NATAUPSKY
PAUL A. STEWART
JOSEPH F. JENNINGS
CRAIG S SUMMERS
ANNEMARIE KAISER
BRENTON R. BABCOCK†
MICHAEL H. TRENHOLM
DIANE M. REED
NANCY WAYS VENSKO
JONATHAN A. BARNEY
RONALD J SCHOENBAUM
RICHARD C GILMORE
JOHN R KING
WILLIAM S REIMUS
CHRISTINE A GRITZMACHER
JOHN P. GIEZENTANNER
ADEEL S AKHTAR
FREDERICK S BERRETTA
THOMAS R ARNO
DAVID N WEISS

CHRISTOPHER A. COLVIN
DANIEL HART
JAMES T. HAGLER
DOUGLAS G. MUEHLHAUSER
LORI L. YAMATO
MOSES MARES
STEPHEN M. LOBBIN
RICHARD KIM
ANN A. BYUN
GLENN R. SMITH
ROBERT F. GAZDZINSKI
FRED C. HERNANDEZ
STACEY R. HALPERN†
R. SCOTT WEIDE
MICHAEL K. FRIEDLAND
DALE C. HUNT
CHAD W. MILLER
JOHN C. WILSON
LEE W. HENDERSON
LENA A. BASILE
JON W. GURKA
DEBORAH S. SHEPHERD
RICHARD E. CAMPBELL

OF COUNSEL
JERRY R. SEILER

JAPANESE PATENT ATTYS
YASUO MURAMATSU†††
KATSUHIRO ARAI

EUROPEAN PATENT ATTY
MARTIN HELLEBRANDT

SCIENTISTS & ENGINEERS
(NON-LAWYERS)

RAIMOND J. SALENIEKS**
RENEE E. CANUSO**
MICHAEL L. FULLER**
NEIL S. BARTFELD**
MICHAEL J. GILLY
HALIT N. YAKUPOGLU
STANLEY W. HUTH
DANIEL E. JOHNSON**
JIAWEI HUANG

Assistant Commissioner for Patents
Washington, D.C. 20231

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

Attorney Docket No.	:	VIRAGE.007A
Applicant(s)	:	Hampapur, et al.
For	:	KEY FRAME SELECTION
Attorney	:	Ray Salenieks
"Express Mail" Mailing Label No.	:	EM365508625US
Date of Deposit	:	June 6, 1997

I hereby certify that the accompanying

Transmittal in Duplicate; Specification in 26 pages; 10 sheets of drawings;
UNSIGNED Declaration and Power of Attorney in 3 pages; Check(s) for Filing
Fee(s); Return Prepaid Postcard

are being deposited with the United States Postal Service "Express Mail Post Office to
Addressee" service under 37 CFR 1.10 on the date indicated above and are addressed to the
Assistant Commissioner for Patents, Washington, D.C. 20231.

Ray Salenieks

JMC-7340:060597

620 NEWPORT CENTER DRIVE
SIXTEENTH FLOOR
NEWPORT BEACH, CALIFORNIA 92660-8016
(714) 760-0404
FAX (714) 760-9502

3801 UNIVERSITY AVENUE
SUITE 710
RIVERSIDE, CALIFORNIA 92501
(909) 781-9231
FAX (909) 781-4507

* A PROFESSIONAL CORPORATION
† ALSO MEMBER OF D.C. BAR
†† ALSO BARRISTER AT LAW (U.K.)
††† ALSO MEMBER OF CA BAR
** U.S. PATENT AGENT

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

ATTENTION: APPLICATION BRANCH

Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): Arun Hampapur, Monika Gorkani, Chiao-Fe Shu and Amarnath Gupta

For: KEY FRAME SELECTION

Enclosed are:

(10) sheet(s) of drawings.

(X) **UNSIGNED** declaration and power of attorney enclosed.

(X) Return prepaid postcard.

CLAIMS AS FILED				
FOR	NUMBER FILED	NUMBER EXTRA	RATE	FEE
Basic Fee			$770	$770
Total Claims	22 - 20 =	2 ×	$22	$ 44
Independent Claims	3 - 3 =	0 ×	$80	$ 0
If application contains any multiple dependent claims(s), then add			$260	$ 0
		TOTAL FILING FEE		$814

(X) The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Account No. 11-1410. A duplicate copy of this sheet is enclosed.

(X) A check in the amount of $814 to cover the filing fee is enclosed.

Ray Salenieks
Registration No. 37,924
Attorney of Record

JMC-7325:060497

DECLARATION AND POWER OF ATTORNEY - USA PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled KEY FRAME SELECTION; the specification of which is attached hereto;

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above;

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56;

I hereby claim the benefit under Title 35, United States Codes § 119(e) of any United States provisional application(s) listed below.

Application No.: 60/019,281   Filing Date: June 7, 1996

POWER OF ATTORNEY: I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith (if this application is assigned, I acknowledge that the appointed individuals do not represent me, and that instead they represent the assignee): Louis J. Knobbe, Registration No. 18,780; Don W. Martens, Registration No. 21,107; Gordon H. Olson, Registration No. 20,319; James B. Bear, Registration No. 25,221; Darrell L. Olson, Registration No. 28,247; William B. Bunker, Registration No. 29,365; William H. Nieman, Registration No. 30,201; Lowell Anderson, Registration No. 30,990; Arthur S. Rose, Registration No. 28,038; James F. Lesniak, Registration No. 25,240; Ned A. Israelsen, Registration No. 29,655; Drew S. Hamilton, Registration No. 29,801; Jerry T. Sewell, Registration No. 31,567; John B. Sganga, Jr., Registration No. 31,302; Edward A. Schlatter, Registration No. 32,297; Gerard von Hoffmann, Registration No. 33,043; Joseph R. Re, Registration No. 31,291; John M. Carson, Registration No. 34,303; Andrew H. Simpson, Registration No. 31,469; Daniel E. Altman, Registration No. 34,115; Anita M. Kirkpatrick, Registration No. 32,617; Ernest A. Beutler, Registration No. 19,901; Vito A. Canuso, Registration No. 35,471; William H. Shreve, Registration No. 35,678; Stephen C. Jensen, Registration No. 35,556; Steven J. Nataupsky, Registration No. 37,688; Michael H. Trenholm, Registration No. 37,743; AnneMarie Kaiser, Registration No. 37,649; Brenton R. Babcock, Registration No. 39,592; Nancy Ways Vensko, Registration No. 36,298; Jonathan A. Barney, Registration No. 34,292; Ronald J. Schoenbaum, Registration No. 38,297; Richard C. Gilmore, Registration No. 37,335; John R. King, Registration No. 34,362; William S. Reimus, Registration No. 38,279; John P. Giezentanner, Registration No. 39,993; Frederick S. Berretta, Registration No. 38,004; Christopher A. Colvin, Registration No. 39,147; Richard Kim, Registration No. 40,046; Glenn R. Smith, Registration No. 38,308; Craig S. Summers,

Registration No. 31,430; Robert F. Gazdzinski, Registration No. 39,990; Yasuo Muramatsu, Registration No. 38,684; Raimond J. Salenieks, Registration No. 37,924; R. Scott Weide, Registration No. 37,755; Renée E. Canuso, Registration No. 36,657; Michael L. Fuller, Registration No. 36,516; Neil S. Bartfeld, Registration No. 39,901; Daniel E. Johnson, Registration No. 37,033; and Katherine W. White, Registration No. 37,470.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issued thereon.

_ _

Full name of first inventor: Arun Hampapur

Inventor's signature ___Unsigned___

Date ___June 6, 1997___

Residence: 209D - 36th Street, San Mateo, CA 94403

Citizenship: **India**

Post Office Address:

Full name of second inventor: Monika Gorkani

Inventor's signature ___Unsigned___

Date ___June 6, 1997___

Residence: 60 Parkridge Drive, San Francisco, CA 94131

Citizenship: **United States of America**

Post Office Address:

Full name of third inventor: Chiao-Fe Shu

Inventor's signature ___*Unsigned*___

Date ___June 6, 1997___

Residence: 302 - 37th Avenue, San Mateo, CA 94403

Citizenship: **Taiwan**

Post Office Address:

Full name of fourth inventor: Amarnath Gupta

Inventor's signature ___*Unsigned*___

Date ___June 6, 1997___

Residence: 1254 Woodside Road, Redwood City, CA 94061

Citizenship: **India**

Post Office Address: 1254 Woodside Road, Redwood City, CA 94061

UNSIGNED DECLARATION AND POWER OF ATTORNEY

Send Correspondence To:
KNOBBE, MARTENS, OLSON & BEAR, LLP
620 Newport Center Drive
Sixteenth Floor
Newport Beach, CA 92660-8016

Direct Telephone Calls To:
John M. Carson

(619) 235-8550

FOR-69P
JMC-7327
060497

KEY FRAME SELECTION

Related Applications

This application claims the benefit of the filing date of U.S. patent application
Serial No. 60/019281, filed June 7, 1996, for "VIRAGE VIDEO: SHOT
SEGMENTATION AND KEY FRAME SELECTION", to Hampapur.

Background of the Invention

Field of the Invention

The present invention relates to video management systems. More specifically,
the invention is directed to a system for automatically processing a video sequence to
extract metadata that provides an adequate visual representation of the video.

Description of the Related Technology

The management of video data is a critical information management problem.
The value of video footage can be effectively utilized only when it can be reused and
repurposed in many different contexts. One of the key requirements to effectively
access video from a large collection is the ability to retrieve video information by
content. Content-based retrieval of video data demands a computer-readable
representation of video. This representation of the original video data is called
metadata. The metadata includes a representation of the visual, audio and semantic
content. In other words, a good representation of a video should effectively capture
the look of the video, its sound and its meaning. An effective representation of the
video captures the essence of the video in as small a representation as possible. Such
representations of the video can be stored in a database. A user trying to access video
from a collection can query the database to perform a content-based search of the
video collection to locate the specific video asset of interest. Figure 1 illustrates a
block diagram of a video database system 100. Such a system is described in
Designing Video Data Management Systems, Arun Hampapur, University of
Michigan, 1995, which is herein incorporated by reference. Video data 102 is input

into a Metadata Extraction module 104. The resultant metadata is stored in a database system 106 by use of an insertion interface 108.

The extraction (104) of metadata from the actual video data 102 is a very tedious process called video logging or manual annotation. Typically this process requires on average labor of eight times the length of the video. What is desired is a system which would automatically process a video so as to extract the metadata from a video sequence of frames that provides a good visual representation of the video.

Some of the terminology used in the description of the invention will now be discussed. This terminology is explained with reference to a set of example images or frames shown in Figure 2. Image one shows a brown building 120 surrounded by a green lawn 122 with a blue sky 124 as a background. Image two shows a brown car 126 on a green lawn 128 with a blue sky 130 as a background. Let us assume that these two frames are taken from adjacent shots in a video. These two frames can be compared based on several different sets of image properties, such as color properties, distribution of color over the image space, structural properties, and so forth. Since each image property represents only one aspect of the complete image, a system for generating an adequate representation by extracting orthogonal properties from the video is needed. The two images in Figure 2 would appear similar in terms of their chromatic properties (both have approximately the same amount of blue, green and brown color's) but would differ significantly in terms of their structural properties (the location of edges, how the edges are distributed and connected to each other, and so forth).

An alternate scenario is where the two images differ in their chromatic properties but are similar in terms of their structural properties. An example of such a scenario occurs when there are two images of the same scene under different lighting conditions. This scenario also occurs when edit effects are introduced during the film or video production process like when a scene fades out to black or fades in from black.

Given any arbitrary video, the process used for generating an adequate visual representation of the video must be able to effectively deal with the situations outlined

in the above discussion. The use of digital video editors in the production process is increasing the fraction of frames which are subjected to digital editing effects. Thus an effective approach to generating adequate visual representations of videos is desired that uses both chromatic and structural measurements from the video.

Several prior attempts at providing an adequate visual representation of the visual content of a video have been made: Arun Hampapur, Designing Video Data Management Systems, The University of Michigan, 1995; Behzad Shahraray, Method and apparatus for detecting abrupt and gradual scene changes in image sequences, AT&T Corp, 32 Avenue of the Americas, New York, NY 10013-2412, 1994, European Patent Application number 066327 A2; Hong Jiang Zhang, Stephen W Smoliar and Jian Hu Wu, A system for locating automatically video segment boundaries and for extracting key-frames, Institute of System Science, Kent Ridge, Singapore 0511, 1995, European Patent Application number 0 690413 A2; and Akio Nagasaka and Yuzuru Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances", Proceedings of the 2nd Working Conference on Visual Database Systems, p.119-133, 1991. Most existing techniques have focused on detecting abrupt and gradual scene transitions in video. However, the more essential problem to be solved is deriving an adequate visual representation of the visual content of the video.

Most of the existing scene transition detection techniques, including Shahraray and Zhang et al., use the following measurements for gradual and abrupt scene transitions: 1) Intensity based difference measurements wherein the difference between two frames from the video which are separated by some time interval "T", is extracted. Typically, the difference measures include pixel difference measures, gray level global histogram measures, local pixel and histogram difference measures, color histogram measures, and so forth. 2) Thresholding of difference measurements wherein the difference measures are thresholded using either a single threshold or multiple thresholds.

However, to generate an adequate visual representation of the visual content of the video, a system is needed wherein the efficacy of the existing techniques is not critically dependent on the threshold or decision criteria used to declare a scene break or scene transition. Using existing techniques, a low value of the threshold would

result in a oversampled representation of the video, whereas, a higher value would result in the loss of information. What is needed is a system wherein the choice of the decision criteria is a non-critical factor.

<div align="center">5</div>

Summary of the Invention

One embodiment of the present invention includes a computer-based system for identifying keyframes or a visual representation of a video by use of a two stage measurement process. Frames from a user-selected video segment or sequence are processed to identify the keyframes. The first stage preferably includes a chromatic difference measurement to identify a potential set of keyframes. To be considered a potential frame, the measurement result must exceed a user-selectable chromatic threshold. The potential set of keyframes is then passed to the second stage which preferably includes a structural difference measurement. If the result of the structural difference measurement then exceeds a user-selectable chromatic threshold, the current frame is identified as a keyframe. The two stage process is then repeated to identify additional keyframes until the end of the video. If a particular frame does not exceed either the first or second threshold, the next frame, after a user-selectable time delta, is processed.

The first stage is preferably computationally cheaper than the second stage. The second stage is more discriminatory since it preferably operates on a smaller set of frames. The keyframing system is extensible to additional stages or measurements as necessary.

In one aspect of the invention, there is a computerized method of extracting a key frame from a video, comprising the steps of providing a reference frame; providing a current frame different from the reference frame; determining a chromatic difference measure between the reference frame and the current frame; determining a structure difference measure between the reference frame and the current frame; and identifying the current frame as a key frame if the chromatic difference measure exceeds a chromatic threshold and the structure difference measure exceeds a structure threshold.

<div align="center">-4-</div>

In another aspect of the invention, there is a computerized method of extracting a key frame from a video having a plurality of frames, the method comprising the steps of providing a reference frame; providing a current frame different from the reference frame; determining a first difference measure between the reference frame and the current frame; determining a second difference measure between the reference frame and the current frame; and identifying the current frame as a key frame if the first difference measure exceeds a first threshold and the second difference measure exceeds a second threshold.

In another aspect of the invention, there is a computerized method of extracting a key frame from a video having a plurality of frames, the method comprising the steps of providing a reference frame; providing a current frame different from the reference frame; determining a structure difference measure between the reference frame and the current frame; and identifying the current frame as a key frame if the structure difference measure exceeds a structure threshold.

Brief Description of the Drawings

Figure 1 is a block diagram showing a video data system wherein the presently preferred key frame system may be utilized;

Figure 2 is a block diagram of two exemplary video frames showing chromatic and structural properties useful in operation of a preferred keyframing system that is a portion of the metadata extraction module shown Figure 1;

Figure 3 is a block diagram of the presently preferred keyframing system;

Figure 4 is a block diagram of frame sequences illustrating operation of the preferred keyframing system of Figure 3;

Figure 5 is a top-level operational flow diagram of the key frame selection system shown in Figure 3;

Figure 6 is a block diagram of the two functions utilized in the "chromatic difference measure" function shown in Figure 5;

Figure 7 is a block diagram of a set of functions, based on edge orientation, utilized in the "structural difference measure" function shown in Figure 5;

Figure 8 is a block diagram of a set of functions, based on edge moments, utilized in the "structural difference measure" function shown in Figure 5;

Figure 9 is a diagram showing a set of video frames at the output of the chromatic difference stage of the keyframing system of Figure 3; and

Figure 10 is a diagram showing a set of video frames at the output of the structural difference stage of the keyframing system of Figure 3.

Detailed Description of the Preferred Embodiment

The following detailed description of the preferred embodiment presents a description of certain specific embodiments of the present invention. However, the present invention can be embodied in a multitude of different ways as defined and covered by the claims. In this description, reference is made to the drawings wherein like parts are designated with like numerals throughout.

For convenience, the discussion of the preferred embodiment will be organized into the following principal sections: Introduction, System Overview, Hierarchical Method of Keyframe Extraction, Keyframing Program, and Measurements Types, Image Processing Procedures, and Results and Summary.

1.0 Introduction

A visual representation of a video is a subset of the images chosen from the video based on some sampling criteria. The keyframing algorithm presented here uses a visual similarity metric to extract a visual representation of the video. The visual representation of the video is defined as the smallest subset of frames that can be chosen from the video which adequately represent the video. The adequacy of the visual representation is controlled by the user through the use of a set of thresholds.

An adequate visual representation of a video is a subset of frames which captures all the visual events in the video without duplicating visually similar frames. According to this definition, a visual representation is not adequate if it misses any visually distinct frames from the video. It is also not adequate if two frames in the representation are not sufficiently distinct.

The visual representation of a video depends on the domain from which the video data is derived. For example, a video from a video conference can be adequately represented by choosing one frame from every shot (a continuous take by a video camera), since each shot would have very little action (e.g., has mostly talking head shots). A video from a football game will need more than one frame per shot for an adequate visual representation, since video shots in football games tend to track the play from one end of the field to the other.

The present invention uses a staged hierarchical approach. In this approach, the decision criteria of the first level can be made less rigid to allow an oversampling. The oversampled set can then be further refined at the second stage to remove redundant representation. In addition, the technique presented uses the structure of the contents of the frame in addition to the intensity distributions. The use of structural information from the image makes the approach less sensitive to intensity changes in the video.

2.0 System Overview

A keyframing system 150 used for extracting the visual representation of the video is shown in Figure 3. A keyframing algorithm that is a portion of the keyframing software 160 operates on Red/Green/Blue (RGB) frame buffers 158 captured from the video. The video can be in a variety of well-known formats, such as analog video 152, MPEG file 154, or D1 format video tape 156. Each of these formats utilizes a suitable video reader or frame grabbers which can be used to digitize or decode the video into a sequence of RGB frame buffers 158. For example, the analog video 152 uses a frame grabber 162, such as Matrox Meteor, the MPEG video 154 uses a MPEG1 decoder 164, such as available from Optivision, and the D1 video 156 uses a D1 reader 166. The keyframing program 160 described below assumes a sequence of RGB frames 158, and a frame number relative to the beginning of the video to be used as a starting frame number. The output of the keyframing program 160 includes a set of keyframe images 172 and corresponding frame numbers 174.

The keyframing system 150 includes a computer 170 that executes the keyframing software 160. The preferred computer is a personal computer having, at

a minimum, an Intel Pentium Pro processor running at 200 MHz, 32 Mb of main memory, and two Gb of mass storage, such as a video-optimized hard drive. The preferred operating software is Windows NT, version 4.0, available from Microsoft. However, other 32-bit operating software systems and comparable processors could

5   be used for running the keyframing program.

3.0 Hierarchical Method of Keyframe Extraction

The method of extracting the visual representation involves a two stage process. The first stage processes the raw video to extract a set of frames which are

10   visually distinct based on the chromatic difference measure and a user supplied chromatic difference threshold. The second stage operates on frames which have been chosen by the first stage. Frames in this stage are compared based on the structure difference measure and a user provided structure difference threshold. Figure 4 shows exemplary sets of frames of the staged hierarchical architecture. The first stage

15   samples frames from a video 200 based on the chromatic activity in the video. The number of frames 202 output by the chromatic difference measurement is proportional to the overall activity in the video 200. A talking head video (e.g., a news anchor person shot) will generate a smaller number of output frames than the video of a sporting event (e.g., a fast break in basketball game).

20   While operating on a typical produced video, such as television feed, the chromatic difference measurement may be tuned to pick up frames during gradual transitions, such as fades, dissolves, wipes and so forth. These frames are typically chromatically different but structurally similar. The redundancy in the output of the chromatic difference based measurement is filtered out by the structural difference

25   measurement, which produces the actual keyframes 204. For example, frames in a fade have the same structure, but are significantly different chromatically due to the fading effect.

Thus, the combination of two or more orthogonal image features in a hierarchical manner provides significant improvement in generating an adequate

30   representation of the video while keeping the computational process simple and efficient. The first feature measurement is selected to be computationally cheaper than

-8-

the second measure.   The second feature measurement is a more discriminatory measurement that extracts more information from a frame than the first measure. The hierarchical method can be extended to "N" stages or measures.

<div align="center">

4.0 <u>Keyframing Program</u>

</div>

This section presents a detailed description of the algorithm for the keyframing program used in this embodiment of the invention. The following list of symbols are used in the description of the algorithm.

4.1   Symbols Used

V	=	Time Indexed Video Sequence (set of RGB frames)
T	=	Current Frame Number
t_b	=	Begin Frame Number
t_e	=	End Frame Number
ΔT	=	Time Increment Factor
i	=	Current Keyframe Number
R	=	Reference Frame
A	=	Active Frame
M_c	=	Chromatic Difference Measure
d_c	=	Chromatic Distance
M_s	=	Structure Difference Measure
d_s	=	Structure Distance
T_c	=	Chromatic Difference Threshold
T_s	=	Structure Difference Threshold
K	=	Keyframe Storage List

4.2   Keyframing Process Steps

Referring to Figure 5, a keyframe selection process 220, which comprises the keyframing software 160 (Figure 3) executed by the computer 170, will now be described. As shown in Figure 3, the input to the program is a sequence of RGB frames, and also includes the initial and final frame numbers of the sequence.

Beginning at a start state 222, process 220 moves to state 224 wherein the Current Video Frame number is initialized to the Initial Frame number of the video sequence $(T = t_b)$, and the Current Keyframe number (i) is initialized to zero. Proceeding to state 226, process 220 sets the Reference Frame to be the Current Video Frame $(R = V(T))$. Continuing at state 228, process 220 updates the video time index $(T = T + \Delta T)$. The time interval, or delta T, is settable by the user of the program, which permits tuning of the keyframe selection process 220. For example, the time interval can be set to advance to the next frame in the sequence, or the time interval can be set to advance four frames in the sequence. The latter case would allow faster processing of the video sequence, but some of the potential keyframes may be missed, which would not provide the best visual representation of the video.

Advancing to a decision state 230, process 220 determines if the end of the video sequence has been reached by checking if Current Frame number is greater than the ending frame number $(T > t_e)$. If so, all the frames in the video have been checked and the keyframe selection process completes at end state 248. If the end of the video sequence has not been reached, as determined at state 230, process 220 proceeds to state 232 wherein the Active Frame is set to be the Current Video Frame $(A = V(T))$. Moving to function 236, process 220 computes the Chromatic Difference Measure between the Active and Reference Frames using the procedure described in section 5.1 below $(d_c = M_c(R, A))$.

Proceeding to a decision state 238, process 220 determines if the chromatic distance derived by function 236 is below the chromatic threshold $(d_c < T_c)$. The chromatic threshold is settable by a user of the keyframing system. If the chromatic distance is below the chromatic threshold, that is, there is not enough chromatic change between the two frames being compared, the Current Frame is not a candidate to be a key frame. Process 220 then moves back to state 228 wherein the next frame to be compared is selected. If the chromatic distance is equal to or greater than the chromatic threshold, the Current Frame is a candidate to be a key frame and corresponds to one of the frames 202 (Figure 4). Process 220 then passes the frame on to the next stage at function 240 wherein the Structure Difference Measure is computed between the Active and Reference Frames using the procedures in section

-10-

5.2 ($d_s = M_s(R, A)$). Note that either the procedure to determine a Structural Difference based on Edge Orientation M_{So} or the procedure to determine a Structural Difference based on Edge Moments M_{Sm} may be used, as determined by the user's needs.

5   Proceeding to a decision state 242, process 220 determines if the structure distance derived by function 240 is below the structure threshold ($d_c < T_s$). The structural threshold is settable by a user of the keyframing system. If the structural distance is below the structural threshold, that is, there is not enough structural change between the two frames being compared, the Current Frame is not a key frame.

10   Process 220 then moves back to state 228 wherein the next frame to be compared is selected. If the structural distance is equal to or greater than the structural threshold, the Current Frame is identified as a key frame and corresponds to one of the frames 204 (Figure 4). Process 220 then proceeds to state 244 and sets the Current Keyframe to the Current Video Frame ($K(i) = V(T)$) to facilitate selection of the reference frame

15   at state 226. Process 220 continues at state 246 and increments the Current Keyframe Number ($i = i + 1$). The keyframe and frame number are preferably stored in an array or list indexed by Current Keyframe Number (i). Process 220 then moves back to state 226 to start the keyframe process again using the new keyframe identified at state 244 as a new Reference Frame. Process 220 continues to process the frames in the

20   video sequence until the end is reached, as determined at state 230.

5.0 Measurements Types

The algorithm described in section 4.2 has two primary image feature extraction processes namely, the chromatic difference measurement and the structural

25   difference measurement. The chromatic measurements filter the video based on the brightness and color differences between the frames. The degree of discrimination provided by any particular specific chromatic measure is bounded due to the fact that these measures rely on the color and intensity distributions. Applying the structural difference metric to the set of frames selected by the chromatic difference metric

30   provides a new dimension along which the frames can be compared. The arrangement

of these two metrics in a hierarchy along with the use of the thresholds allows the efficient generation of adequate visual representations of the video.

5.1 Chromatic Difference Measurement: $M_c(R, A)$ (236, Figure 5)

The chromatic difference measurement operates on a pair of frames (RGB buffers) and computes the chromatic distance between the frames. Chromatic difference measurements cover a wide range of measurements, such as luminance pixel differences, color pixel differences, local intensity histogram differences, global intensity histogram differences and so forth. In this embodiment of the invention, a gray level intensity histogram-based chromatic difference measurement is utilized.

5.1.1 Chromatic Difference Measurement based on Intensity Histograms

This measurement uses the gray level intensity histogram of the two frames. This is a measure of how the intensities vary in the frame. The histogram of the reference frame is compared to the histogram of the active frame using the χ^2 metric. The χ^2 distance is used as the chromatic difference between the reference and active frames. The steps in the algorithm are discussed below. The functions used in the chromatic difference measurement (236) and the functional interrelationship are shown in Figure 6.

Step 1:   Compute the intensity histogram of reference frame H_R using procedure in section 6.2.

Step 2:   Compute the intensity histogram of the active frame H_A using procedure in section 6.2.

Step 3:   Compute the difference of the histograms using the procedure in section 6.8.

Step 4:   Set the chromatic difference to be the χ^2 distance.

X^2_H =   The histogram difference measurement

$H_A(i)$ =   n bit gray scale histogram of the Active Frame

$H_R(i)$ =   n bit gray scale histogram of the Reference Frame

N =   is the number of gray levels

5.2   Structure Difference Measurement: M_S (240, Figure 5)

This measurement operates on two RGB frames and computes the structural distance between the frames. The structure difference measurement includes any measurement which compares images based on the structure (edge) content of the image. In this embodiment of the invention, edge orientation histogram difference and edge moment difference are utilized as two types of structure difference measurement techniques. Either type can be used as the structural difference measurement 240.

5.2.1   Structural Difference based on Edge Orientation M_{So}

This measurement computes the structural difference between the reference and active frames by measuring the χ^2 difference between the edge orientation histograms of the two images. The edge orientation histogram captures the global structure of the image. It captures the dominant directions in which the major edges in the image are distributed. The difference measure is generated by comparing the two edge orientation histograms using the χ^2 difference metric. The steps in the algorithm are discussed below. The functions used in this edge orientation type measurement (240) and the functional interrelationship are shown in Figure 7.

Step 1:   Let E_R be the edge mask for reference image R using procedure in section 6.6

Step 2:   Let E_A be the edge mask for active image A using procedure in section 6.6

Step 3:   Let $G_{R\theta}$ be the gradient orientation image of the reference image computed using the procedure in section 6.4

Step 4:   Let $G_{A\theta}$ be the gradient orientation image of the active image computed using the procedure in section 6.4

Step 5:   Let $H_{R\theta}$ be the edge orientation histogram computed based on E_R and $G_{R\theta}$ using procedure in section 6.7

Step 5:   Let $H_{A\theta}$ be the edge orientation histogram computed based on E_A and $G_{A\theta}$ using procedure in section 6.7

Step 6:   Compute the χ^2 difference between the orientation histograms using $H_{R\theta}$
            and $H_{A\theta}$

Step 7:   Set the structure difference to be the χ^2 distance.

5   5.2.2   Structural Difference based on Edge Moments M_{Sm}

            The moments of the edge image are a statistical measure of the spatial
distribution of edges in the image. In this embodiment, the first five moments of the
edge image are utilized to represent the structure of the image. The moments are
normalized. The moments of the reference and active images are compared by

10   computing an energy difference between the two sets of moments. The algorithm for
comparing the moment based structural difference is presented below. The functions
used in this edge moments type measurement (240') and the functional
interrelationship are shown in Figure 8.

15   Step 1:   Let E_R be an edge image of the reference frame generated using the
                procedure in section 6.6.

     Step 2:   Let E_R be an edge image of the active frame generated using the
                procedure in section 6.6.

     Step 3:   Let M, N be the number of moments to be computed in the X and Y

20   directions.

     Step 4:   Let m_R be the moment set for the reference image computed using the
                procedure in section 6.9

     Step 5:   Let m_A be the moment set for the active image computed using the
                procedure in section 6.9

25   Step 6:   Let d_s be the difference in the moments of m_R, m_A computed using the
                procedure in section 6.11

                        6.0 Image Processing Procedures
            The following procedures are used in computing the Measurements from the frames.

30   The procedures described in here are used by practitioners in the field of computer
vision. Most of these algorithms can be found in text books dealing with computer

vision. Specifically most of the information used here in has been derived from Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995, incorporated herein by reference.

6.1 Symbols Used

H	=	Histogram of the frame
i	=	Gray Level Intensity at a Pixel
F	=	RGB video frame
F_r	=	Red channel of F
F_g	=	Green channel of F
F_b	=	Blue channel of F
δx	=	Index into the frame
δx	=	Index increment
Δy	=	Index increment
y	=	Index into the frame
X	=	Width of the frame in pixels
Y	=	Height of the frame in pixels

6.2 Gray Level Intensity Histogram Computation

This process uses a color (RGB) image and generates the luminance or brightness histogram of the image.

Step 1:   Set the image indices to 0

$x = 0, y = 0$

Step 2:   Increment the image index

$x = x + \delta x$

Step 3:   If $x > X$ go to Step 10

Step 4:   Set

$y = 0$

Step 5:   Increment the image index

$y = y + \delta y$

-15-

Step 6:   If $y > Y$ go to Step 2

Step 7:   Compute the Intensity value at the pixel.

$I = 0.114 \times F_r(x,y) + 0.587 \times F_g(x,y) + 0.299 \times F_b(x,y)$

Step 8:   Increment the corresponding histogram bin

$H(I) = H(I) + 1$

Step 9:   Go to Step 5

Step 10:   End

6.3   Luminance Image Computation

This computation uses a color image (RGB) and converts it into a gray scale image by combining the individual color bands of the image. The constants used in Step 7 can be found in Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995.

Step 1:   Set the image indices to 0

$x = 0, y = 0$

Step 2:   Increment the image index

$x = x + \delta x$

Step 3:   If $x > X$ go to Step 9

Step 4:   Set

$y = 0$

Step 5:   Increment the image index

$y = y + \delta y$

Step 6:   If $y > Y$ go to Step 2

Step 7:   Compute the Intensity value at the pixel.

$I(x,y) = 0.114 \times F_r(x,y) + 0.587 \times F_g(x,y) + 0.299 \times F_b(x,y)$

Step 8:   Go to Step 5

Step 9:   End

6.4   Gradient Orientation Image Computation

This process acts on an RGB image to produce an image where each pixel in the image represents the direction or angle of the gradient (step 4) at that pixel. This is an intermediate step in the computation of edge orientation histogram.

Step 1: Let I be the intensity image generated from RGB buffer using the procedure in section 6.3.

Step 2: Let the G_x be x gradient image generated using the Sobel edge mask $M_y(i,j)$ (see page 147, Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

$G_X(x,y) = I(x,y) * M_y(i,j)$

Step 3: Let the G_Y be y gradient image generated using the Sobel edge mask $M_Y(i,j)$ (see page 147, Ramesh, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

$G_Y(x,y) = I(x,y) * M_x(i,j)$

Step 4: Let G_θ be the gradient orientation image.

$$G_\theta(t, x, y) = tan^{-1}(\frac{G_Y(t, x, y)}{G_X(t, x, y)}).$$

6.5   Gradient Magnitude Image Computation

This process acts on an RGB buffer to produce an image where each pixel represents the magnitude of the gradient (step 4) at that point. This is an intermediate step in the computation of an edge image.

Step 1: Let I be the intensity image generated from RGB buffer using the procedure in section 6.3.

Step 2: Let the G_x be x gradient image generated using the Sobel edge mask $M_X(i,j)$ (see page 147, Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

$G_X(x,y) = I(x,y) * M_y(i,j)$

Step 3: Let the G_Y be y gradient image generated using the Sobel edge mask $M_Y(i,j)$ (see page 147, Ramesh Jain and Rangachar Kasturi and Brian G Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

$G_Y(x,y) = I(x,y) * M_x(i,j)$

5 Step 4: Let G_M be the gradient magnitude image.

$$G_M(t, x, y) = \sqrt{G_X^2 + G_Y^2}$$

6.6 Edge Image Computation

An edge image is an image which outlines only the significant edges in the
10 source image. A pixel in an image is marked as a significant edge if the gradient magnitude at that point exceeds a preset edge threshold. The value of the edge threshold is experimentally chosen. There are several automatic techniques for selecting thresholds discussed in literature (Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

15

Step 1: Let G_M be the gradient magnitude image computed using the procedure in section 6.5

Step 2: Let T_e be a predetermined edge threshold.

Step 3: Let E be the edge image generated by thresholding the gradient
20 magnitude image (see page 143, Ramesh Jain and Rangachar Kasturi and Brian G. Schunck, Introduction to Machine Vision, McGraw Hill, 1995).

6.7 Orientation Histogram Computation

25 The orientation histogram captures the distribution of edge orientations in the image. The following are the steps in the orientation histogram computation procedure. This procedure operates on an edge image and a gradient orientation image to generate an orientation histogram.

30 Let E be an edge image generated using the procedure in section 6.6.

Let G_θ be the gradient magnitude image generated using the procedure in section 6.4.

Step 1: Set the image indices to 0

$x = 0, y = 0$

Step 2: Increment the image index

$x = x + \delta x$

Step 3: If $x > X$ go to Step 11

Step 4: Set

$y = 0$

Step 5: Increment the image index

$y = y + \delta y$

Step 6: If $y > Y$ go to Step 2

Step 7: If the current pixel is not a valid edge pixel.

$E(x,y) \neq$ Valid Edge Pixel Go to Step 10

Step 8: Let $\theta = G_\theta(x,y)$

Step 9: Increment the corresponding histogram bin

$H(\theta) = H(\theta) + 1$

Step 10: Go to Step 5

Step 11: End

6.8 χ^2 Histogram Difference Computation

This is a specific type of histogram comparison. This technique does a bin by bin differencing of the two histograms and normalizes the difference by the sum of the corresponding bins in the histogram. The normalization makes the differencing less sensitive to small changes in the histogram. The following is the procedure for computing the χ^2 difference of two histograms H_1 and H_2.

$$X_H^2 = \sum_{i=0}^{N} \frac{|H_1(i) - H_2(i)|^2}{H_1(i) + H_2(i)} \tag{3}$$

$N = $ *is the number of bins*

6.9   Edge Moment Set Computation

The moments are a statistical measure of the edges properties in the image. The lower order moments capture the gross structure of the edges (like the centroid of edges) and the higher order moments capture the finer variations in the edge structure (like corners, highly curved edges etc). The following is the algorithm for computing the moments.

Step 1:   Let M be the number of moments to be computed in the X direction.

Step 2:   Let N be the number of moments to be computed in the Y direction.

Step 3:   Set the image indices to 0

$m = -1, n = -1$

Step 4:   Increment the index

$m = m + 1$

Step 5:   If $m > M$ go to Step 10

Step 6:   Set

$n = -1$

Step 7:   Increment the index

$n = n + 1$

Step 8:   If $n > N$ go to Step 4

Step 9:   Compute the moment $M(m,n)$ using the procedure outlined in section 6.10.

Step 10:   End

6.10   Edge Moment Value Computation

This procedure computes the $(m,n)^{th}$ moment of the edge image. This moment is computed based on the centroid of the edge image. The moments are normalized. The following formulae can be used to compute the moments.

6.11   Edge Moment Difference

$$v(m,n) = \frac{e_v(m,n)}{e_n(m,n)} \tag{4}$$

$$e_v(m,n) = \sum_{x=0}^{X}\sum_{y=0}^{Y}(x - \bar{x})^m \times (y - \bar{y})^n \times E(x,y) \tag{5}$$

$$\bar{x} = \frac{\sum_{x=0}^{X}\sum_{y=0}^{Y} x \times E(x,y)}{X * Y} \tag{6}$$

$$\bar{y} = \frac{\sum_{x=0}^{X}\sum_{y=0}^{Y} y \times E(x,y)}{X * Y} \tag{7}$$

$$e_n(m,n) = \sum_{x=0}^{X}\sum_{y=0}^{Y}\left\| (x - \bar{x})^{(m+n)} \times E(x,y) \right\| + \sum_{x=0}^{X}\sum_{y=0}^{Y}\left\| (y - \bar{y})^{(m+n)} \times E(x,y) \right\| \tag{8}$$

There are several different techniques for computing the structure difference between frames using edge moments. In this embodiment, the structure difference is computed by finding the root mean square difference between the moment sets using equation 9.

5

$$d = \sum_{m=0}^{M}\sum_{n=0}^{N}\sqrt{(m_1(m,n) - m_2(m,n))^2} \tag{9}$$

7.0 Illustrative Results and Summary

10 The invention presented in the above sections has been applied to a wide variety of video sequences. Figures 9 and 10 show the output of the chromatic and structural stages. The images in Figures 9 and 10 are frames extracted from a video sequence, the number assigned to each image is the frame number of the image in the video sequence. The exemplary video sequence starts at frame number 1790 and ends

15 at frame number 2389 for a total of 600 frames. The video has been digitized at thirty frames per second. Thus two images, which have frame numbers thirty frames apart, are spaced one second apart in the video. The images in these figures are arranged from left to right and top to bottom in order the increasing order of time.

The output of the chromatic difference measurement (Figure 9) has twenty-four frames, and clearly, some of these frames are structurally similar. The chromatic difference measure selects frames 1802, 1804, 1833 as they are part of a fade in sequence where there are significant changes in the chromatic measurements. Frames 1921, 1937, 1950 are selected due to the fact that there is large object motion in the frame as it is a extreme close up shot. Frames 2146 to 2201 are selected due to the high degree of specular reflection in a close up shot. Frames 2280 - 2312 are selected due to the large object motion in an extreme close up shot.

The output of the structural difference measurement (Figure 10) has fourteen frames. These frames are clearly structurally different and comprise an adequate visual representation of the video. The structural difference measurement eliminates the structurally similar frames.

The results discussed in this section clearly illustrate the benefits and strengths of the present invention. The approach clearly recognizes the limitations of relying completely on chromatic metrics and applies a more sophisticated measurement to overcome these limitations. The computational expense of the algorithm is kept small by using the hierarchical approach which allows the more expensive computations to be applied to a smaller set of frames. The structural computation is more discriminatory than the chromatic computation.

While the above detailed description has shown, described, and pointed out the fundamental novel features of the invention as applied to various embodiments, it will be understood that various omissions and substitutions and changes in the form and details of the system illustrated may be made by those skilled in the art, without departing from the intent of the invention.

WHAT IS CLAIMED IS:

1.   A computerized method of extracting a key frame from a video, comprising the steps of:

        a) providing a reference frame;

        b) providing a current frame different from the reference frame;

        c) determining a chromatic difference measure between the reference frame and the current frame;

        d) determining a structure difference measure between the reference frame and the current frame; and

        e) identifying the current frame as a key frame if the chromatic difference measure exceeds a chromatic threshold and the structure difference measure exceeds a structure threshold.

2.   The method defined in Claim 1, additionally comprising the step of setting the current frame to be the reference frame if a key frame is identified.

3.   The method defined in Claim 1, additionally comprising the step of repeating steps c-e for a new current frame until the end of the video is reached.

4.   The method defined in Claim 3, wherein the new current frame is selected to be at a predetermined time interval after the current frame.

5.   The method defined in Claim 4, wherein the predetermined time interval is user-selectable.

6.   The method defined in Claim 1, wherein the value of the chromatic threshold and the value of the structure threshold are each user-selectable.

7.   The method defined in Claim 1, wherein the step of determining the structure difference measure is performed only if the chromatic difference measure exceeds the chromatic threshold.

8.	A computerized method of extracting a key frame from a video having a plurality of frames, the method comprising the steps of:

a) providing a reference frame;

b) providing a current frame different from the reference frame;

c) determining a first difference measure between the reference frame and the current frame;

d) determining a second difference measure between the reference frame and the current frame; and

e) identifying the current frame as a key frame if the first difference measure exceeds a first threshold and the second difference measure exceeds a second threshold.

9.	The method defined in Claim 8, additionally comprising the step of setting the current frame to be the reference frame if a key frame is identified.

10.	The method defined in Claim 8, wherein the first difference measure is orthogonal to the second difference measure.

11.	The method defined in Claim 8, additionally comprising the step of repeating steps c-e for a new current frame until the end of the video is reached.

12.	The method defined in Claim 11, wherein the new current frame is selected to be at a predetermined time interval after the current frame.

13.	The method defined in Claim 8, wherein the value of the first threshold and the value of the second threshold are each user-selectable.

14.	The method defined in Claim 8, wherein the step of determining the second difference measure is performed only if the first difference measure exceeds the first threshold.

15. The method defined in Claim 8, wherein the second difference measure is computationally more expensive than the first difference measure.

16. The method defined in Claim 8, wherein the second difference measure extracts more information than the first difference measure.

17. The method defined in Claim 8, additionally comprising the step of determining a third difference measure between the reference frame and the current frame, and wherein the identifying step identifies the current frame as the key frame if the third difference measure exceeds a third threshold.

18. A computerized method of extracting a key frame from a video having a plurality of frames, the method comprising the steps of:

    a) providing a reference frame;

    b) providing a current frame different from the reference frame;

    c) determining a structure difference measure between the reference frame and the current frame; and

    d) identifying the current frame as a key frame if the structure difference measure exceeds a structure threshold.

19. The method defined in Claim 18, additionally comprising the step of setting the current frame to be the reference frame if a key frame is identified.

20. The method defined in Claim 18, additionally comprising the step of repeating steps c and d for a new current frame until the end of the video is reached.

21. The method defined in Claim 20, wherein the new current frame is selected to be at a predetermined time interval after the current frame.

22. The method defined in Claim 18, wherein the value of the structure threshold is user selectable.

KEY FRAME SELECTION

Abstract of the Disclosure

A system and method that processes video to extract a keyframe-based adequate visual representation. The method utilizes a hierarchical processing technique. The first stage in the hierarchy extracts a chromatic difference metric from a pair of video frames. An initial set of frames is chosen based on the chromatic metric and a threshold. A structural difference measurement is extracted from this initial set of frames. A second threshold is used to select key frames from the initial set. The first and second thresholds are user selectable. The output of this process is the visual representation. The method is extensible to any number of metrics and any number of levels.
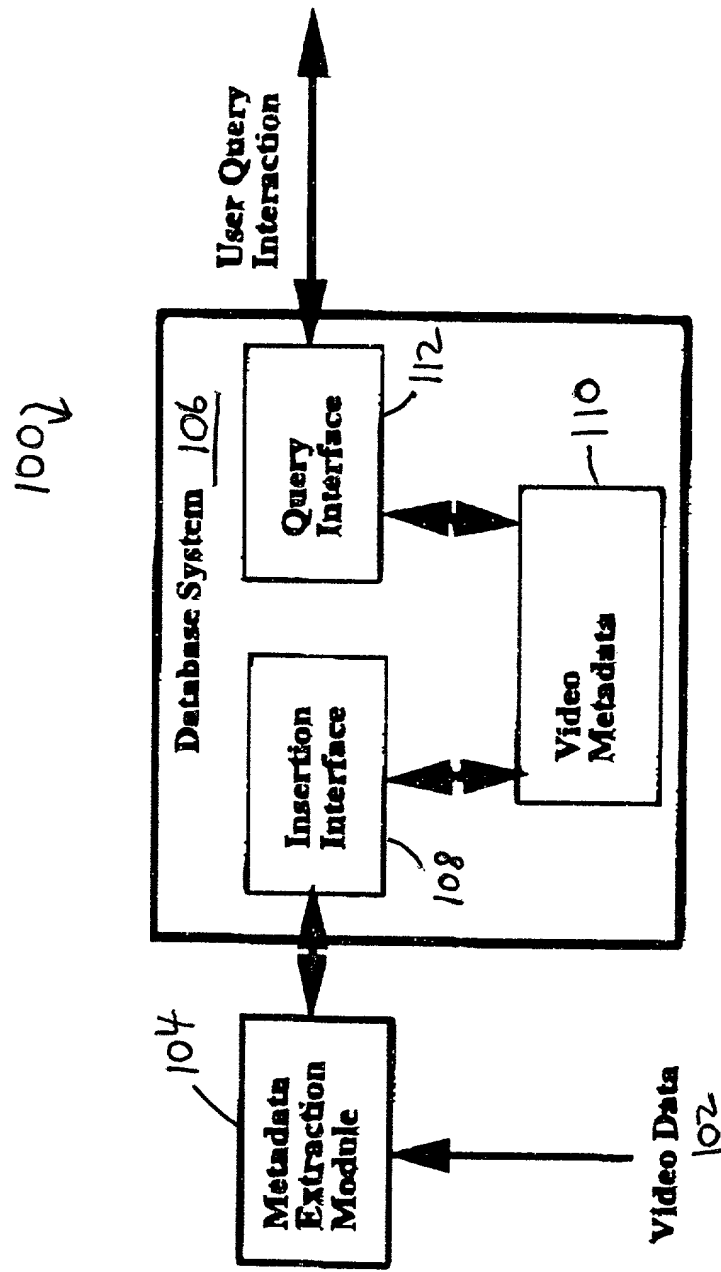
5

10
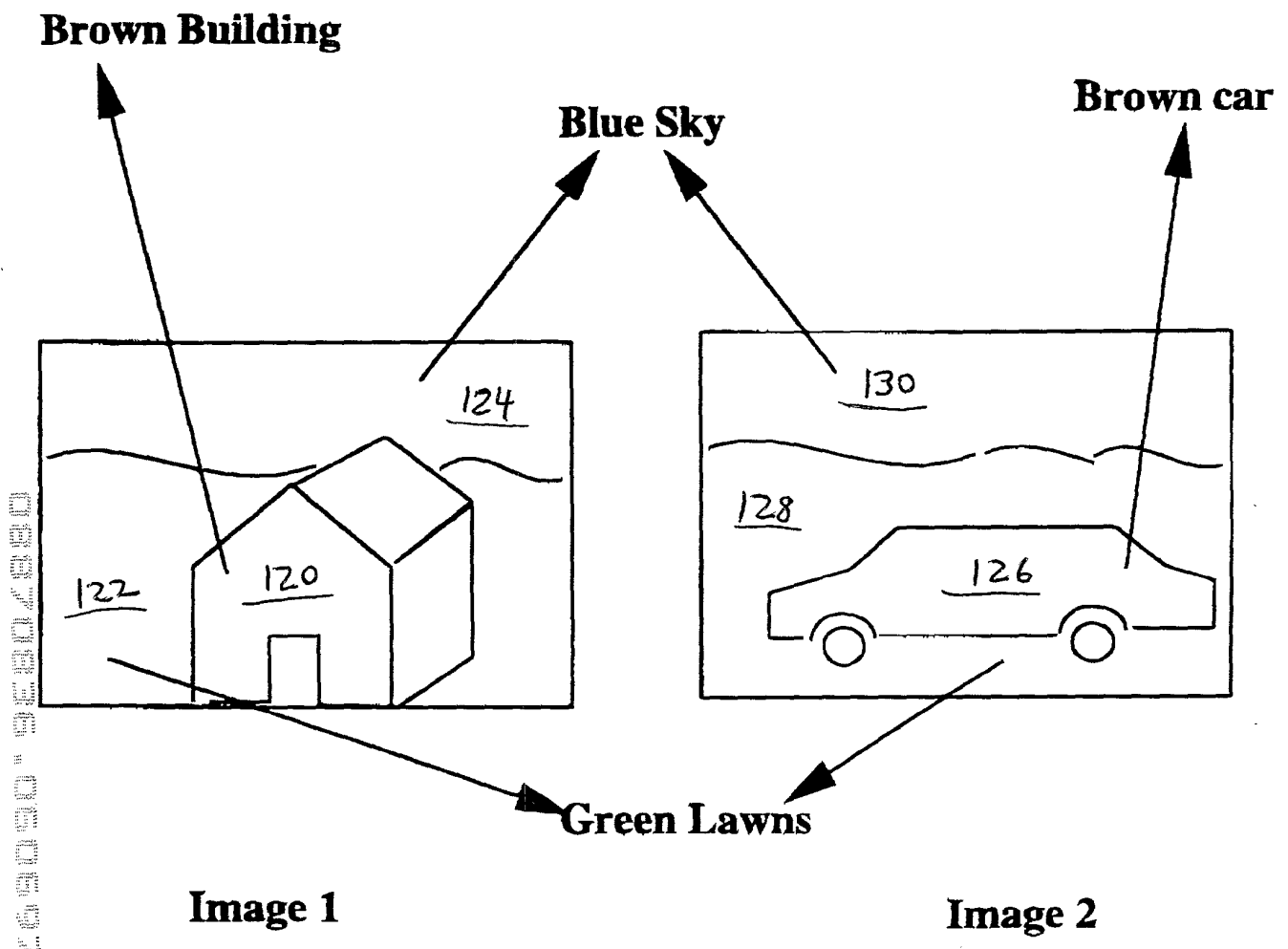
15

Figure 1: A Video Database System

Brown Building

Blue Sky

Brown car

124

130

128

122   120

126

Green Lawns

Image 1

Image 2

Figure 2 Examples to illustrate the Chromatic and Structural Properties of Images

Video Readers

150

Analog
Video

152

Matrox
Meteor

Frame
Grabber

162

Mpeg
File

154

Mpeg 1
Decoder

Optivision

164

RGB
Frame

158

Pentium Pro
200 Mhz
PC

170

Keyframing
Software

160

Keyframe
Image

172

Frame
Number

174

D1
Tape

156

D1
Reader

166
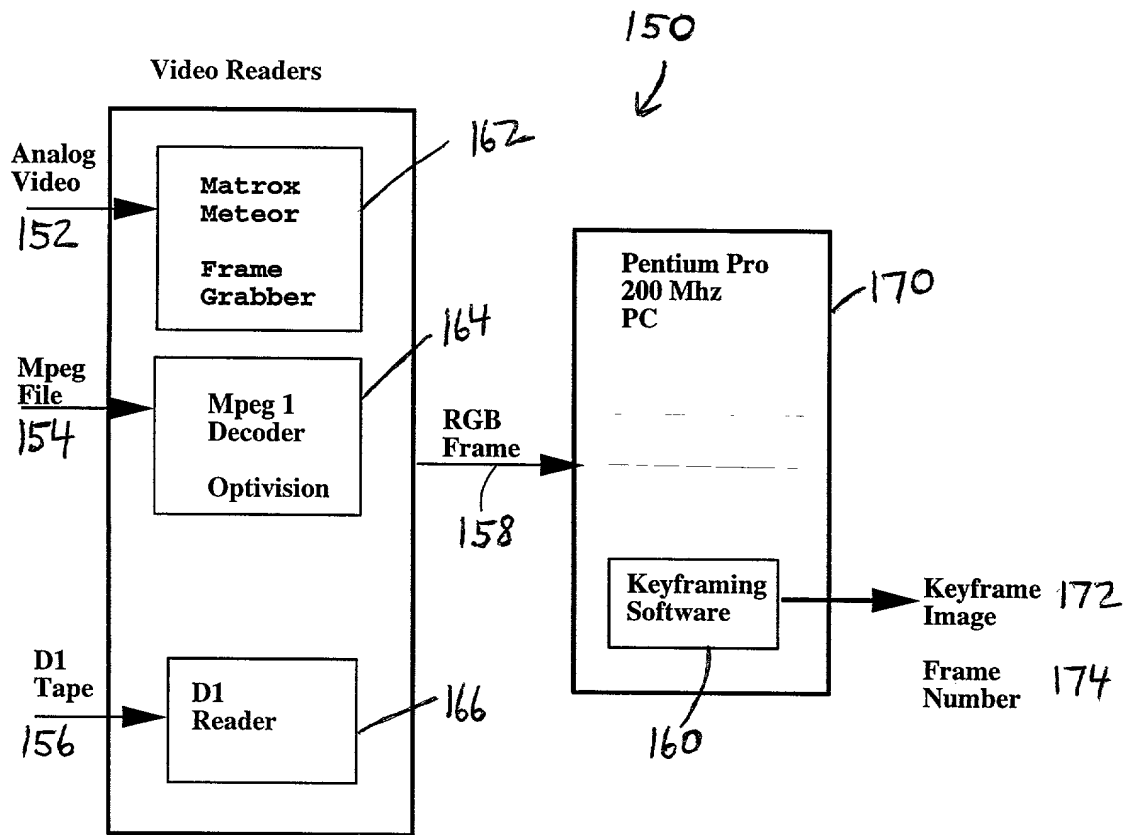
Figure 3 The Keyframing System

Figure 4

222 — START   220

224 — INITIALIZE CURRENT FRAME NUMBER TO BEGINNING FRAME NUMBER AND CURRENT KEYFRAME NUMBER TO ZERO
[T = tb, i = 0]

226 — SET THE REFERENCE FRAME TO BE THE CURRENT VIDEO FRAME
[R=V(T)]

228 — UPDATE THE VIDEO TIME INDEX
[T = T + delta T]

230 — T>END FRAME NUMBER(te)?   YES

NO   232

232 — SET ACTIVE FRAME TO BE CURRENT VIDEO FRAME
[A = V(T)]

246 — INCREMENT CURRENT KEYFRAME NUMBER
[i = i + 1]

COMPUTE THE CHROMATIC DIFFERENCE MEASURE BETWEEN THE ACTIVE AND REFERENCE FRAMES
[dc = Mc(R,A)]

236

CHROMATIC DIFFERENCE BELOW CHROMATIC THRESHOLD
[dc < Tc]?   238

YES

NO

COMPUTE STRUCTURE DIFFERENCE MEASURE BETWEEN ACTIVE AND REFERENCE FRAMES
[ds = Ms(R,A)]

240

248 — END

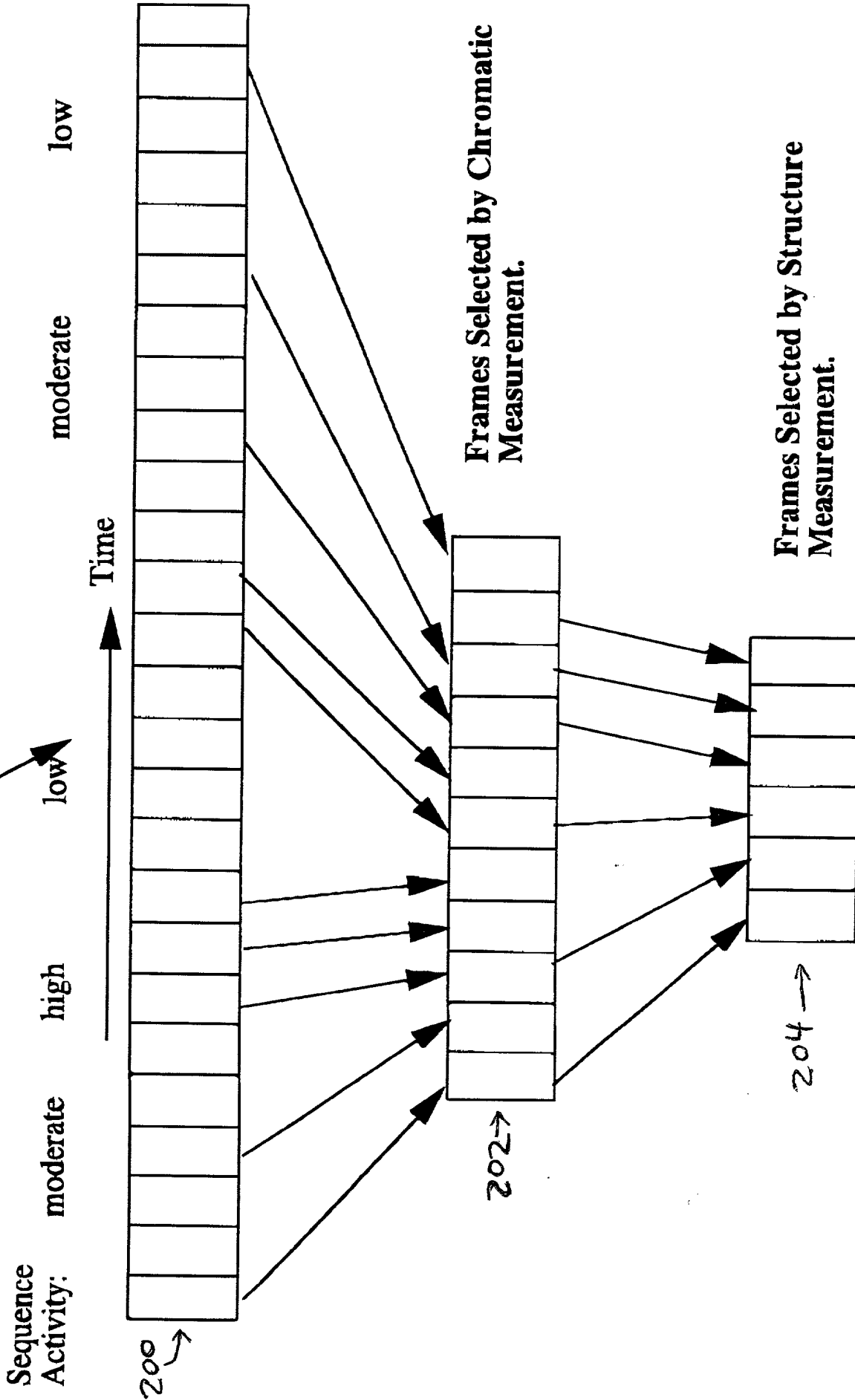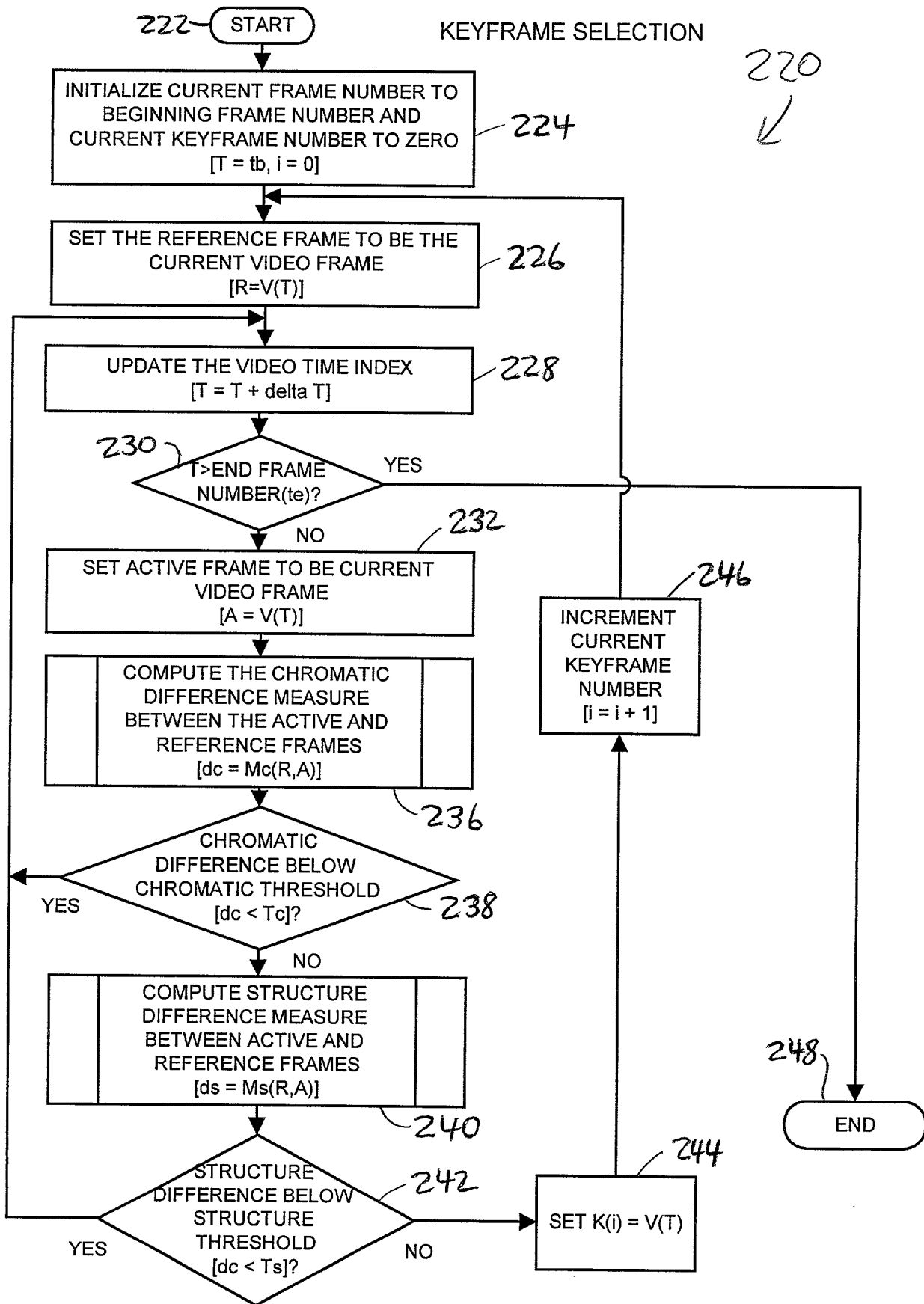STRUCTURE DIFFERENCE BELOW STRUCTURE THRESHOLD
[dc < Ts]?   242

YES   NO

244 — SET K(i) = V(T)

FIGURE 5

FIGURE 6

FIGURE 7

```
STRUCTURAL
DIFFERENCE MEASURE
BASED ON EDGE              240'
MOMENTS
(SECTION 5.2.2)


EDGE IMAGE          EDGE              EDGE
(6.6)          MOMENT SET         MOMENT
                  (6.9)          DIFFERENCE
                                    (6.11)


GRADIENT             EDGE
MAGNITUDE           MOMENT
IMAGE               VALUE
(6.5)               (6.10)


LUMINANCE
IMAGE
(6.3)
```

FIGURE 8

1802   1804   1833   1873

1921   1937   1950   1956

2043   2091   2146   2155

2162   2181   2186   2201

2227   2240   2280   2291

2298   2306   2312   2328

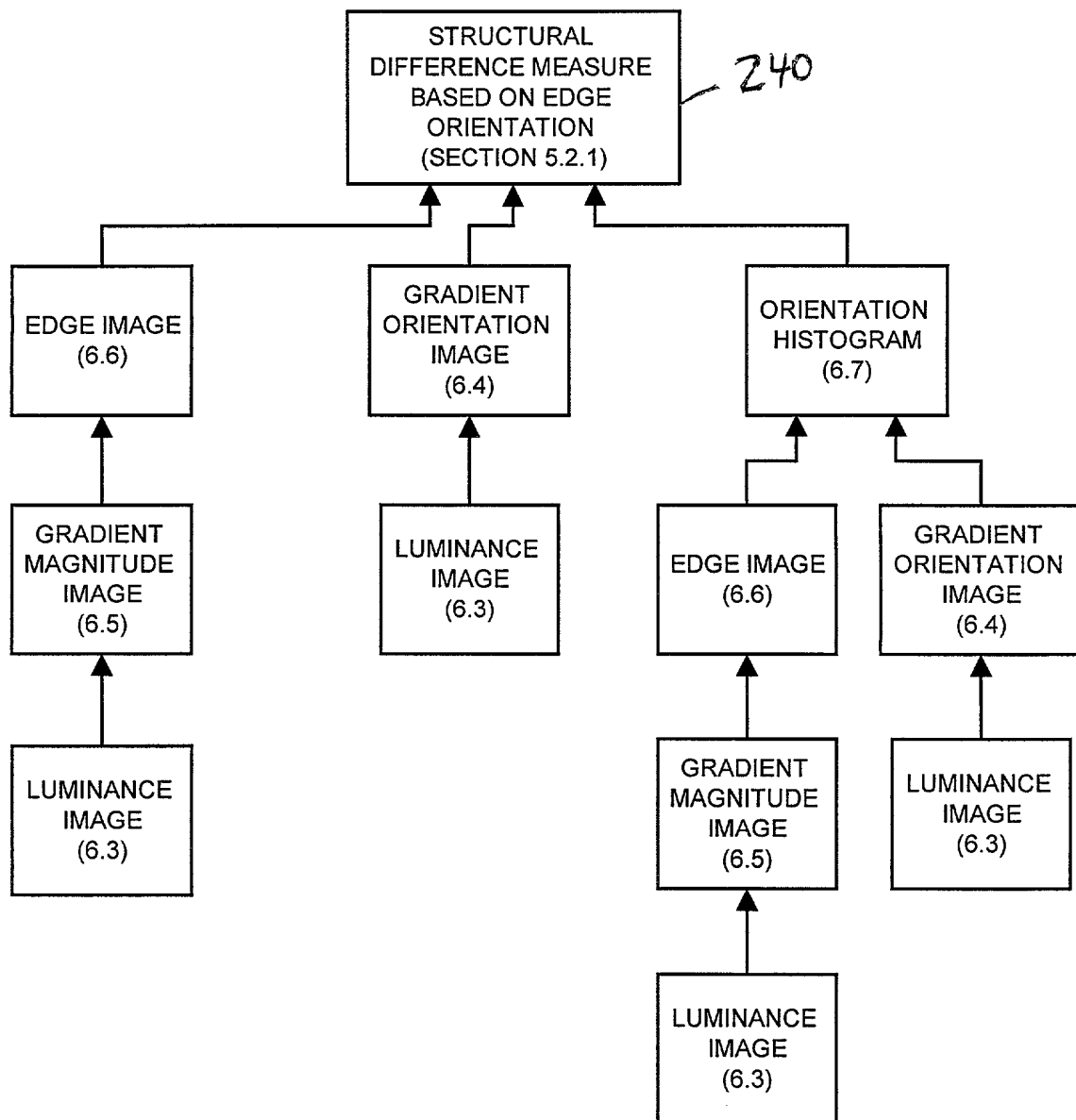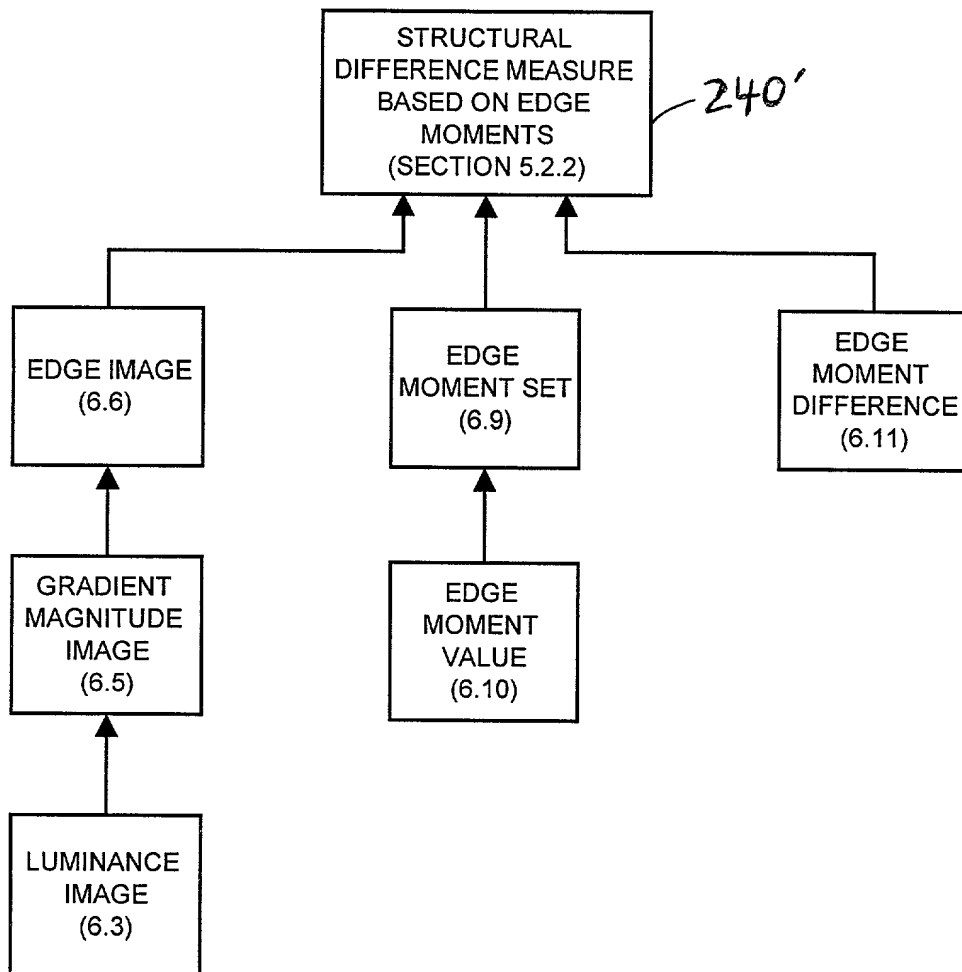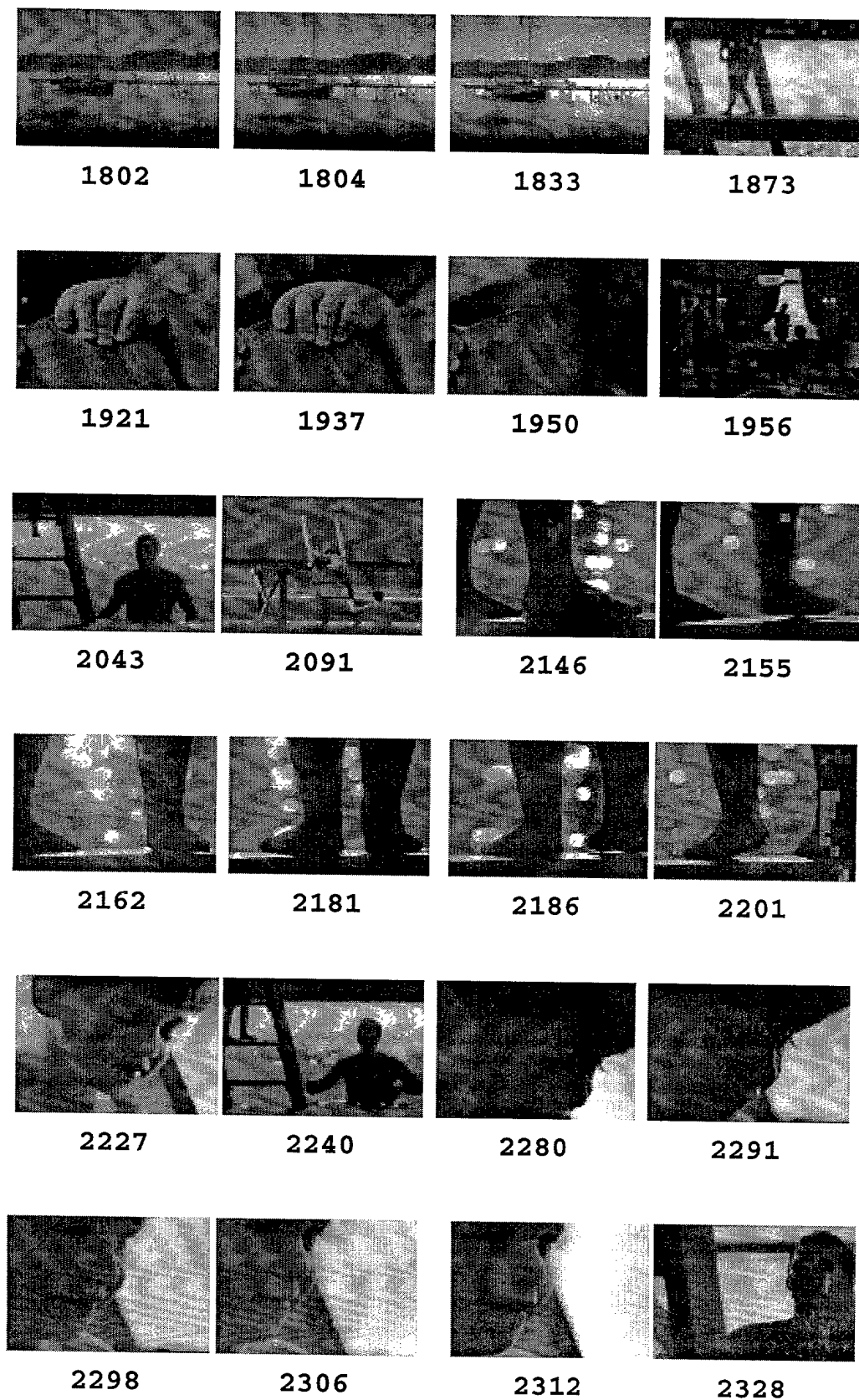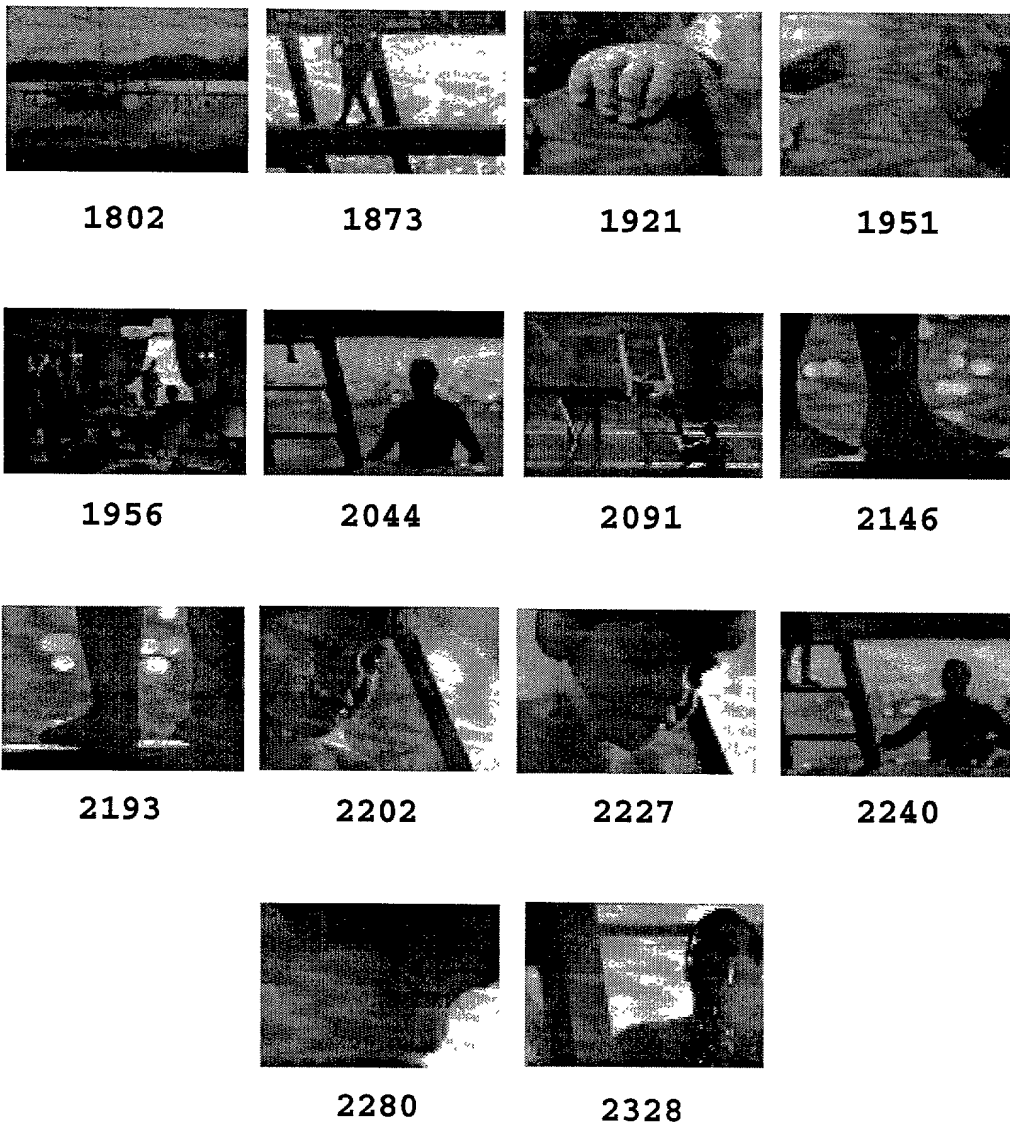## Output of Stage 1: Chromatic Difference (236)

Figure 9 The Results of Applying the invention to 600 frames of a Video sequence

1802   1873   1921   1951

1956   2044   2091   2146

2193   2202   2227   2240

2280   2328

Output of Stage 2: Structural Difference (240)

Figure 10 The Results of Applying the invention to 600 frames of a Video sequence